Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Kuehn HS, Boisson B, Cunningham-Rundles C, et al. Loss of B cells in patients with heterozygous mutations in IKAROS. N Engl J Med 2016;374:1032-43. DOI: 10.1056/NEJMoa1512234

Supplementary Appendix

Supplement to: Kuehn HS, Boisson B, et al. Loss of B Cells in Patients with Heterozygous Mutations in IKAROS

Contents

Loss of B Cells in Patients with Heterozygous Mutations in IKAROS

Hye Sun Kuehn*, Ph.D., Bertrand Boisson*, Ph.D., Charlotte Cunningham-Rundles, M.D., Ph.D., Janine Reichenbach, M.D., Asbjørg Stray-Pedersen, M.D., Ph.D., Erwin W. Gelfand, M.D., Patrick Maffucci, B.A., Keith R. Pierce, B.S., Jordan K. Abbott, M.D., Karl V. Voelkerding, M.D., Sarah T. South, Ph.D., Nancy H. Augustine, B.A., Jeana S. Bush, M.D., William K. Dolen, M.D., Betty B. Wray, M.D., Yuval Itan, Ph.D., Aurelie Cobat, M.D., Ph.D., Hanne Sørmo Sorte, M.S., Sundar Ganesan, Ph.D., Seraina Prader, M.D., Thomas B. Martins, M.S., Monica G. Lawrence, M.D., Jordan S. Orange, M.D., Ph.D., Katherine R. Calvo, M.D., Ph.D., Julie E. Niemela, M.S., Jean-Laurent Casanova, M.D., Ph.D., Thomas A. Fleisher, M.D., Harry R. Hill, M.D., Attila Kumánovics, M.D., Mary Ellen Conley, M.D.,#§, Sergio D. Rosenzweig, M.D., Ph.D.#

* Drs. Kuehn and Boisson contributed equally, # Drs. Conley and Rosenzweig contributed equally, § Corresponding author

Department of Laboratory Medicine, NIHCC (H.S.K., J.E.N., K.R.C., T.A.F., S.D.R.) and Primary Immunodeficiency Clinic, NIAID (S.D.R), National Institutes of Health, Bethesda, MD; St. Giles Laboratory of Human Genetics of Infectious Diseases, The Rockefeller University, New York City, NY (B.B., Y.I., A.C., J.-L.C., M.E.C.); Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Inserm U1163 and Paris Descartes University, Imagine Institute, Paris, France (A.C., J.-L. C.), and Howard Hughes Medical Institute, New York, NY, and Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Inserm U1163 and Pediatric Hematology-Immunology Unit, Necker Hospital for Sick Children, Paris, France, and Paris Descartes University, Imagine Institute, Paris, France (J.-L.C); Department of Medicine and the Immunology Institute, Icahn School of Medicine at Mount Sinai, NY (C.C.-R., P.M.); Division of Immunology, University Children's Hospital Zurich and Children's Research Center (J.R., S.P.) and University Zurich, Switzerland (J.R); Center for Human Immunobiology of Texas Children's Hospital/Department of Pediatrics, Baylor College of Medicine, Houston, TX (A.S.-P., J.S.O), Baylor-Hopkins Center for Mendelian Genomics of the Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX and Norwegian Unit for National Newborn Screening, Oslo University Hospital, Oslo, Norway (A.S.-P.); Division of Allergy and Immunology, Department of Pediatrics, National Jewish Health, Denver, CO (J.K.A., E.W.G.); University of Tennessee College of Medicine, Memphis, TN (K.R.P.); Department of Pathology (K.V.V., S.T.S., N.H.A., T.B.M, H.R.H., A.K.) and Department of Pediatrics and Medicine (H.R.H), University of Utah School of Medicine and ARUP Institute for Clinical and Experimental Pathology, ARUP Laboratories, Salt Lake City, UT; Division of Allergy-Immunology and Pediatric Rheumatology, Department of Pediatrics, Medical College of Georgia at Georgia Regents University, Augusta, GA (J.S.B.,W.K.D., B.B.W.); Department of Medical Genetics, Oslo University Hospital and University of Oslo, Norway (H.S.S.); Biological Imaging Section, Research Technologies Branch, NIAID, NIH (S.G.), Division of Asthma, Allergy and Immunology, Department of Medicine, University of Virginia, Charlottesville, VA (M.G.L.)

Supplementary Clinical Information

Family A

A1 is a Hispanic female who was evaluated for immunodeficiency at 10 years of age because of a history of recurrent respiratory tract infections and a recent episode of pneumococcal meningitis. At 2 weeks of age she was hospitalized with fever and a seizure. Starting at 1 year of age, she had 10-15 episodes of bronchitis or otitis per year. In the year preceding evaluation, she had 2 episodes of pneumonia, one requiring hospitalization. One month preceding evaluation she was hospitalized with meningoencephalitis. Spinal fluid was purulent and cultures grew *S. pneumoniae*. Sequelae included temporary right-sided hearing loss and mild intellectual impairment. Laboratory evaluation at 10 years of age revealed mild thrombocytopenia, hypogammaglobulinemia, absent antibodies to vaccine antigens and markedly reduced numbers of B cells (Table 1). She was started on gammaglobulin replacement but compliance was intermittent. She was hospitalized at 20 years of age for pneumonia and treated as an outpatient for pneumonia several times. At 24 years of age, high resolution CT demonstrated mild airway disease but no bronchiectasis or nodules. A2, the son of A1, was evaluated for immunodeficiency at 6 years of age because of recurrent upper respiratory infections, acute otitis media and sinusitis. At age 3 years he developed immune thrombocytopenia (ITP; platelets 46,000/μl). He did not respond to steroid therapy but Anti-D Ig therapy controlled his ITP for 2 years. At 5 years of age, he presented with a second episode of ITP and was successfully treated with 4 doses of rituximab, (platelets 388,000/μL). Before rituximab treatment, his serum IgG concentration was low at 431mg/dL (IgA, IgM and B cells were not determined) and he was placed on gammaglobulin replacement. At age 6 years his IgG was normal, IgA was low as was IgM. B cells in peripheral blood were also very low (CD19 0.3%, 12/μL); a residual effect of

rituximab treatment could not be completely ruled out. A2 has done well, with no hospitalizations and no recurrences of ITP for the last 4 years.

Family B

B5 is a Swiss female who was evaluated for immunodeficiency at 3 years of age because of recurrent otitis, cough, dental caries and apthous ulcers (Table 1). Serum immunoglobulins were reported to be decreased and antibody titers to pneumococcus were low despite repeated vaccinations. She continued to have frequent episodes of otitis and bronchitis and was started on gammaglobulin therapy at 15 years of age, which resulted in a decrease in the incidence of otitis and bronchitis. At 16 years of age she had an episode of diarrhea due to *Giardia lamblia*, which responded to treatment. She has had recurrent herpes labialis since age 20 and she had an episode of diarrhea associated with *Blastocystis hominis* at 23 years of age. B6, the brother of B5 was recognized to have hypogammaglobulinemia at 1 year of age after multiple episodes of otitis, bronchitis and apthous ulcers (Table 1). He was hospitalized for pneumonia at 2 and 6 years of age, and again at 9 years of age, when he had pneumococcal sepsis and pneumonia. He was started on gammaglobulin therapy at 9 years of age. He has done well since then but continues to have recurrent apthous ulcers. B1, the paternal grandfather of B5 and B6 had a long history of recurrent respiratory tract infections and was sent to a TB sanitarium at 12 years of age because of a chronic cough. He had a cerebral hemorrhage at 30 years of age and a cerebral infarction at 52 years of age. He was evaluated for immunodeficiency at 71 years of age and given the diagnosis of CVID (Table 1). He was treated with gammaglobulin for less than a year and died of pneumonia at 74 years of age. The DNA

sample from this individual was obtained at the time of his diagnosis with CVID. B2, B3, B4 and B7 were recognized to have immunodeficiency during the family studies.

B2, the son of B1 and the father of B5 and B6, had recurrent otitis and bronchitis as a small child. He had a severe pneumonia requiring partial pneumonectomy at 6 years of age (Table 1). He had the onset of recurrent purulent skin infections at 20 years of age and he continues to have recurrent bronchitis. B3, the daughter of B1, had a basal skull fracture as the result of a scooter accident at 6 years of age and developed meningitis with hearing loss. She had a second episode of meningitis at 10 years of age but has done well otherwise. B4, the daughter of B1, has had recurrent otitis and bronchitis since early childhood. B7, the daughter of B3, developed B cell acute lymphoblastic leukemia (ALL at 3 years of age and died of relapsed ALL at 5 years of age. The leukemia was described as having strong expression of CD19 and CD10, weak expression of CD20 and CD22 with some cells positive for CD34 and HLA-DR. Five of 16 metaphases were positive for an interstitial deletion of the region around Chromosome 13q.14. The DNA sample from B7 was obtained from bone marrow taken when she was diagnosed as having ALL.

Family C

C1 is a European-American who was evaluated for immunodeficiency at 32 years of age. She was reportedly a healthy child, although she had 2 episodes of "walking pneumonia" at less than 13 years of age. Her spleen was removed after an automobile accident when she was 27 years old. Three years later, she was hospitalized with pneumococcal meningitis, with full recovery after adequate antibiotic treatment. At 32 years of age, she had two separate episodes of pneumococcal sepsis requiring admissions to the ICU. After the second septic episode, immune evaluation demonstrated hypogammaglobulinemia affecting IgG, IgA and IgM (Table 1) and

"very few B cells in peripheral blood". She was started on gammaglobulin replacement. Prophylactic antibiotics were added when she was 40 years old and she has not had major infections since that time.

Patient C2 is the eldest daughter of C1. She was a healthy child but given her mother's history was immunologically evaluated at age 9 months (Table 1). Because her serum immunoglobulins were below normal for age with low B cell numbers, she was placed on gammaglobulin replacement. By 4 years of age her B cells dropped below 2%. Because she remained asymptomatic, a trial off gammaglobulin replacement was attempted at age 5 years. The patient maintained adequate IgG and IgA, but not IgM levels for 8 years, but was unable to mount protective responses to tetanus toxoid, diphtheria toxoid, varicella and rubella after proper and repetitive immunizations. During the same period her B cell numbers continued to decline. At age 13 years of age, after one episode of pneumonia, the patient was restarted on gammaglobulin replacement with good response. Patient C3 is the second daughter of C1. During her neonatal period she had a short NICU admission because of mild bronchiolitis; she also developed 4 episodes of acute otitis media during her first 18 months of life. At that age, per her mother's request, she had her first immune evaluation showing slightly low IgG and IgM, while IgA was normal (Table 1). B cells were present although low for age and protective tetanus toxoid and diphtheria toxoid titers were also detected. For the following 3 years she had repeated, although not severe, upper respiratory infections. At 4 years of age she had pneumonia and was placed on gammaglobulin replacement. Previous to initiating replacement, her laboratory tests showed hypogammaglobulinemia, no response to recall immunizations with tetanus toxoid and diphtheria toxoid, and B cells still present in peripheral blood although at low number per age. During follow up, B cells continued to decline reaching <2% by the age of 11y. On gammaglobulin replacement the patient has had no recurrence of infections for a follow up

period of almost 10 years.

Family D

D2 is a European-American who was reportedly well until she developed pneumonia at 9 years of age. At that time she had severe hypogammaglobulinemia of the three major isotypes and her medical records indicate that she had "normal B cells" (Table 1). No further studies or specific prophylactic therapy were initiated. She was started on gammaglobulin replacement at 12 years of age after a second pneumonia; but discontinued therapy at 16 years of age. At 23 years of age, chronic cough prompted a new evaluation that revealed severe hypogammaglobulinemia and markedly decreased B cells. For the last 2 years she has been on IgG replacement therapy and has not had significant infections. D1 is the mother of D2. She was identified during family studies analyzing the segregation of mutation c.551G>A, p.R184Q in family D. She is a self-reported healthy 50-year-old woman. Her immune evaluation showed mild hypogammaglobulinemia with normal B cells.

Family E

E1 is a Norwegian male who was treated with penicillin in the first month of life for facial skin infections. He had the onset of recurrent respiratory infections in early childhood with multiple episodes of pneumonia, two of which were associated with sepsis (blood cultures positive for *S. pneumoniae* and *H. influenzae*, respectively). At 3 years of age he developed a *S. pneumonia* right hip osteoarthritis. Immune evaluation at 4 years of age revealed severe hypogammaglobulinemia and low B cells (Table 1) and he was started on gammaglobulin replacement therapy. This significantly reduced the number and severity of his infections He continues to have

sporadic upper respiratory infections and diarrhea but he is otherwise well. His son, E2, is 6 years old and reportedly healthy. He was identified during family studies analyzing the segregation of the deletion in family E.

Family F

F2 is a European-American who was evaluated for immunodeficiency at 29 years of age because of recurrent pneumonias and severe sinusitis. Her serum immunoglobulins at that time were undetectable (Table 1). She was started on suboptimal doses of intramuscular gammaglobulin (2 mL/month) but continued to have severe sinus infections and multiple episodes of pneumonia. At 52 years of age she had surgery for maxillary sinusitis and was treated with multiple antibiotics. Intravenous gammaglobulin was recommended but she was unable to secure funding for this. Her serum immunoglobulins at that time were undetectable (Table 1). At 56 years of age, she suffered pneumonias and had increasing diarrhea and weight loss. *C. difficile* toxin was discovered in her stool and she was treated with vancomycin. At 57 years of age, she had not had gammaglobulin therapy for the previous 4 years, and her IgG, IgA and IgM were still undetectable. After vaccination, she had no detectable antibody to 14 pneumococcal serotypes. X-rays obtained at that time revealed clear lung fields and normal cardiac size. She was placed on 400 mg/kg of IV gammaglobulin replacement every month and dramatically improved. During the following years, and while on intermittent gammaglobulin replacement, she had 2-4 sinus infections per year, but suffered from no invasive pneumonias, chronic diarrhea or other significant problems. Currently at 70 years of age, she is on 30 grams of IV gammaglobulin replacement every four weeks and has had occasional sinusitis and one episode of vaginitis but otherwise is doing well.

F1, the sister of F2, was first investigated as part of the evaluation of F2's extended family. At

that time, she was 57 years. Her IgG and IgM were low with normal IgA (Table 1). She had very low B cells, normal CD3 cells, and normal CD4/8 ratio. She has refused gammaglobulin replacement and was doing well except for 2-3 sinus infections per year with no other infections.

Patient F3, the 37-year-old son of F1, suffered frequent upper respiratory infections, as a child but had no ear infections or chronic diarrhea. At the 21 years of age, his serum immunoglobulins were extremely low, B cells were <2% while T and NK cells were normal (Table 1). He was started on IV gammaglobulin replacement at 400 mg/kg. At his present age of 40 years he is on regular gammaglobulin replacement, he has 1-2 episodes of sinusitis per year but has no other infections.

Patient F4, the son of F1 suffered recurrent upper respiratory infections since he was a teenager with no documented pneumonias, hospitalizations, sinusitis, or diarrhea. At 24 years of age his serum immunoglobulins were extremely low (Table 1). He was reevaluated at 25 years of age and his immunoglobulins were still extremely low; B cells were also very low, while T and NK cells were normal. Chest X-rays and pulmonary function tests were normal. He was placed on 400 mg/kg of IV gammaglobulin replacement every 4 weeks and has done well except for periodic sinusitis and cough treated with oral antibiotics. His children all remain entirely asymptomatic.

Patient F5 was first studied at 28 years of age as part of the evaluation of the extended family (Table 1). At that time she was essentially healthy, as she reported no pneumonias, sinusitis, draining ear infection, or chronic diarrhea. Her three major serum immunoglobulins were low. She refused gammaglobulin replacement at that time. When contacted in 2015 at 31 years of age, she reported a history of a walking pneumonia, but no diarrhea, sinusitis, or other infections were reported. She still refuses gammaglobulin replacement therapy.

F6 is the son of F2. He had one pneumonia as a child and was noted to have low

immunoglobulins at 19 years of age. He has continued to have 3-5 episodes of sinusitis per year but he has not had significant gastrointestinal complications or additional pneumonias. His three major serum immunoglobulin levels were low when enrolled in the family study at 42 years of age (Table 1). He has refused gammaglobulin replacement.

F12, the son of F6, developed B-cell acute lymphocytic leukemia at 5 years of age after presenting with bruising and joint pain (Table 1). His white blood cell count at that time was 22,000/μl and the leukemia was characterized as being positive for CD19, CD10, CD22, CD24, HLA-DR, CD34 and TdT with normal cytogenetics. The leukemia was successfully treated with chemotherapy. At 6 years of age, he developed fever and a left cheek nodule as well as left upper thigh nodules and had cultures positive for *Mycobacterium chelonae*. He was treated for several months with antibacterial therapy but continued to have additional skin nodules. Because of persistent neutropenia and bone marrow suppression, he was treated with bone marrow transplant from his HLA-matched brother, F13. At 13 years of age, his IgG, IgA, and IgM were low. He has not suffered serious infections since that time and is currently not on immunoglobulin therapy at 17 years of age.

F7, F8, F9, F10, F11, F13 have remained asymptomatic up until July of 2015 with none having pneumonia, sinusitis, or chronic diarrhea in spite of low levels of IgG, IgA, and/or IgM.

Supplementary Materials and Methods

Patients and samples

All patients or their guardians provided informed consent in accordance with the Declaration of Helsinki under institutional review boards. Blood from healthy donors and patients was obtained under approved protocols, which also allow for the collection and use of patients' family history and pedigrees for publication. All procedures were based on standard of care and established clinical guidelines were followed.

Whole exome sequencing and bioinformatics

For family A, B and D, massively parallel sequencing was performed. Genomic DNA extracted from the patient's peripheral blood cells was sheared with a Covaris S2 Ultrasonicator. An adapter-ligated library was prepared with the Paired-End Sample Prep kit V1 (Illumina). Exome capture was performed with the SureSelect Human All Exon kit (Agilent Technologies). Paired-end sequencing was performed on a HiSeq 2500, generating 100-base reads. We used BWA aligner [1] to align sequences with the human genome reference sequence (hg19 build). Downstream processing was carried out with the Genome analysis toolkit (GATK) [2], SAMtools [3] and Picard Tools (http://picard.sourceforge.net). Substitution and indel calls were identified with a GATK Unified Genotyper and a GATK Indel GenotyperV2, respectively. All calls with a read coverage ≤2x and a Phred-scaled SNP quality of ≤20 were filtered out. All the variants were annotated with the GATK Genomic Annotator. We used several genomic databases: Exome Aggregation Consortium (ExAC), Cambridge, MA (exac.broadinstitute.org), 1000 genome project (www.1000genomes.org) and dbSNP

(www.ncbi.nlm.nih.gov/SNP/). CADD scores were calculated from (http://cadd.gs.washington.edu/score).

For family C, genomic DNA was submitted to Otogenetics for whole-exome capture [Agilent V4 (51 Mbp)] and next-generation sequencing on an Illumina HiSeq2000. The DNAnexus interface was used for alignment of the Illumina reads to the hg19 human reference genome and for discovery and genotyping of single-nucleotide polymorphisms and insertions or deletions. To prioritize variant calls, we implemented the ANNOVAR functional annotation package, filtering the output by gene–amino acid annotation, functional prediction scores, nucleotide conservation scores and allele frequencies according to the NCBI dbSNP database (build 137), The 1000 Genomes Project (2012 April release) and the NHLBI GO Exome Sequencing Project (ESP6500). Only nonsynonymous novel variants or variants with population frequencies less than 1% were considered further. Exomic variants were finally prioritized by clinical correlation and analysis of protein structure and relevant protein motifs.

Sanger sequencing

Selected next-generation sequencing results were confirmed by Sanger sequencing. Specifically, *IKZF1* coding exons were amplified by PCR with exon-specific oligonucleotide primers and GoTaq Hot Start Polymerase (Promega) (primer sequences and cycling conditions available upon request). Purified PCR products were directly sequenced using BigDye Terminators (version 1.1) and were analyzed on a 3130xL Genetic Analyzer (Applied Biosystems).

Detection of intragenic deletion on *IKZF1*

Genomic DNA from the proband in family E was analyzed by whole exome sequencing as previously described.[4] Copy number variants were predicted from the whole-exome file using ExCopyDepth and the predicted deletion affecting *IKZF1* was confirmed using a custom high-resolution exon-tiling array, exaCGH.[4] Probes for aCGH were prepared from genomic DNA according to Agilent protocol v.6.3. Agilent Genomic Workbench (v7.0) was used for analysis of exaCGH. PCR to identify deletion breakpoints was performed using AmpliTaq Gold® 360 (Applied Biosystems) and primers spanning the region of the deletion. The purified, amplified product was further Sanger sequenced using BigDye terminators (version 3.1) and analyzed on a 3730 sequencer (Applied Biosystems) to yield the exact breakpoint of the deletion (primers available upon request). [5-8]

Detection of deletion on Chromosome 7

4.7 Mb deletion on the short arm of chromosome 7 (7p12.3-p12.1; shown in red in Figure 1B) testing was done by array comparative genomic hybridization (aCGH) and/or by Multiplex Ligation-Dependent Probe Amplification (MLPA) from MRC Holland, and Molecular Inversion Probe (MIP) panel. Among the 11 genes included in the 7p12.3-p12.1; *ABCA13, CDC14C, VWC2, ZPBP, C7orf7, IKZF1, FIGNL1, DDC, GRB10, COBL, POM121L12,* only *IKZF1* is selectively expressed in hematopoietic cells; moreover, IKAROS is expressed at all stage of lymphoid development, from HSC through to mature B and T cells, as well as in natural killer (NK) and thymic dendritic antigen presenting cells[9-11].

Data analysis

The data from Family A were analyzed based on the assumption that the mutation in A1 was a de novo event. Sequence data from Families B and D were screened for mutations in IKZF1 based on the findings in Family A. The data from Family C and F were independently analyzed based on the hypothesis that the affected family members would share a rare or unreported genetic variant. The data from Family E were analyzed with the knowledge that mutations in *IKZF1* had been identified in the other families.

Plasmid preparation

Human IKAROS family zinc finger 1 (IKAROS) (*IKZF1*; NM_006060) DNA were synthesized (GeneCopoeia and GeneScript) and subcloned into the mammalian expression vector pcDNA3-HA or CMV driven N-terminal DDK-Myc-flag tagged expression vectors (pCMV6-AC, Origene). Indicated mutants for the *IKZF1* were generated based on the site directed mutagenesis protocol using Quick Change II kit (Agilent) or AccuPrime Pfx DNA Polymerase, followed by DpnI treatment (Life technologies).

Confocal microscopy

NIH3T3 cells ($0.8\text{-}1\times10^5$) were seeded onto cover slips in 6 well plates. The next day, cells were transfected with indicated plasmid using Turbofect (Thermo Scientific) or Effectene (Qiagen) according to the manufacturer's instruction. The cells were washed twice in PBS, fixed for 10 min in 4% paraformaldehyde and permeablized for 15 min in 0.1% Triton X-100 in PBS at RT. The cells were then incubated for 30 min in blocking buffer (PBS with 10% FBS and 0.1% Triton X-100) then incubated for 2 hours with mouse anti-HA monoclonal antibody (Covance), and/or rabbit anti-Flag polyclonal

affinity antibody (Sigma)). The cells were then washed with PBS and incubated for 1 hour with either Alexa Fluore 488 (green color) or Alexa Fluore 568 (red color)-conjugated secondary antibodies in blocking buffer. Cells were washed, mounted on slides using VECTASHIELD Mounting Medium (Vector Laboratories). Images were collected on a Leica SP5 inverted confocal microscope with a 63x oil immersion objective NA 1.4 (Leica Microsystems, Buffalo Grove, IL). Post processing and image analysis were performed using, Huygens (SVI imaging, Nederland) Imaris software (Bitplane Inc., South Windsor, CT).

Western blots

Expression vectors (pCMV6-AC-Myc-DDK) were transfected into HEK-293 cells by Lipofectamine LTX (Invitrogen). Nuclear extracts were obtained with NE-PER nuclear and cytoplasmic extraction reagent (Thermo Scientific) 19 hours after transfection. Proteins were separated by SDS-PAGE and transferred to PVDF membranes. The membranes were incubated with antibodies to anti-DDK (Flag. Origene), Lamin A/C (Santa Cruz), or GAPDH (Santa Cruz) and visualized by chemiluminescence using species-specific HRP-linked antibodies (GE Healthcare).

EMSA assay

Nuclear extracts were incubated 10 min at room temperature in 20mM HEPES, 100mM KCl, 10% glycerol with 100nM final ^{32}P-labeled probes in presence of poly-dIdC (100ng/µl). DNA-protein complexes were separated on 5% acrylamide gel with 0.5x TBE. The probes were derived from the regulatory regions of IK-bs4

(Forward: 5'-gtcaTGACAGGGAATACACATTCCCAAAAGC, Reverse: 5'-gtcaGCTTTTGGGAATGTGTATTCCCTGTCA) or Hs-2mer-bs (Forward: 5'-gtcaATCCATAGGGAAAATGTTTTCCCTGACATC, Reverse: 5'-gtcaGATGTCAGGGAAAACATTTTCCCTATGGAT)[12]. Images have been obtained using the Amersham Biosciences Typhoon 9400 Variable Imager and quantify with ImageQuant V.5.0.

Bone marrow staining

Bone marrow flow cytometric analysis was performed on a FACSCanto II flow cytometer and immunohistochemistry was performed on bone marrow core biopsy sections using the Ventana automated staining system as previously described.[13] For the flow cytometric analysis, 500,000 total events were captured per bone marrow sample. Cells depicted are gated on all $CD34^+$ cells and all $CD19^+$ cells in bone marrow aspirate from a healthy control marrow (37,000 gated events) and patient B2 (14,000 gated events). The following antibodies (clone) were used: CD34 (8G12), CD19 (SJ25C1), TdT (E17-1519).

TdT mediated additions in VDJ recombination of Ig heavy chains

Total RNA was extracted from frozen peripheral blood lymphocytes from four healthy adult controls, 2 patients with mutations in BTK (R255X and L452P) and less than 0.2% $CD19^+$ B cells in the peripheral circulation, and B5 and C1 using an RNeasy Mini kit (QIAGEN Inc). cDNA was prepared with random hexamers and SuperScript III First-Strand (Life Technologies Corporation). The VDJ segment of immunoglobulin heavy chain transcripts belonging to the VH3 family was amplified using the following primers:

GGGGTACCATGGAGTTTGGGCTGAG (forward) and GGAATTCTCACAGGAGACGAG (reverse). PCR conditions were 5 minutes at 95 ^{o}C, followed by 35 cycles of 95 ^{o}C melting for 45 seconds, 60 ^{o}C annealing for 30 seconds, and 72 ^{o}C extension for 30 seconds with a final extension of 5 minutes at 72 C. PCR products were cloned into a TA vector (TOPO TA Cloning Kit for Sequencing, Life Technologies Corporation, Norwalk, CT, USA) and sequenced using M13 vector primers. The VDJ sequenced was analyzed using IMGT website tools which indicate the N additions at both the V to D and D to J junctions. Between 8 and 32 unique sequences were analyzed for each control and for each patient.

Evaluation of cell proliferation

Human PBMCs were isolated by density-gradient centrifugation with Ficoll-Paque PLUS (GE Healthcare), washed twice in phosphate buffered saline (PBS) and cultured in complete RPMI-1640 medium containing 10% FBS (FBS), Penicillin-Streptomycin-Glutamine (Life technology) at 37°C in a humidified 5% CO_2 incubator. PBMCs, either freshly isolated or thawed from liquid nitrogen−stored samples, were incubated with CellTrace violet Cell Proliferation Kit (1μM; Invitrogen). After 20 minutes, 10 volumes of RPMI/10% fetal bovine serum (FBS) were added, and the cells were centrifuged and washed twice more with RPMI/10% FBS. A total of 1×10^5 cells were seeded into 96-well plates, stimulated with anti-CD3 antibody and anti-CD28 antibody (1 μg/mL each, eBioscience) as indicated in the figures, and cultured for 3 days. Cells were stained with fluorochrome-conjugated CD4 and CD8 antibodies (BD Biosciences) for 30 minutes at 4°C (dark). Cells were washed with PBS twice, and cells were acquired and analyzed by flow cytometry (Becton Dickinson FACSCanto II) and FlowJo software (TreeStar).

Apoptosis measurement

Total PBMCs were stimulated with anti-CD3 and CD28 (1μg/mL) for 3 days, and then cultured cells with IL-2 (10ng/ml). At day 11, cells were aliquoted in duplicate into 96 well plates and stimulated with Apo1.3 (1 μg/ml; ENZO Life Sciences) in the presence of protein A (1 μg/ml; Sigma), or 5μM staurosporine (Sigma).

After 20 h, cells were incubated with propidium iodide (Life Technology), and live cells (PI negative) were counted by flow cytometry (BD FACSCanto II; BD Biosciences) by constant time acquisition. The percentage of cell loss was calculated according to the following formula: (number of live cells without APO-1-3 treatment − number of live cells with APO-1-3 treatment/number of live cells without APO-1-3 treatment) × 100.

IKAROS intracellular staining

For IKAROS intracellular staining (family A, B, and C), PBMCs were fixed and permeabilized after surface staining ($CD3^+$ and $CD19^+$) using FoxP3 staining kit (eBioscience) following the manufacturer's instructions, and then incubated with anti-IKAROS antibodies (BD, clone R32-1149) on ice for 30 min. All samples were acquired and analyzed by flow cytometry (Becton Dickinson FACSCanto II) and FlowJo software (TreeStar). For family D (bottom panel), goat polyclonal anti-IKAROS antibodies or normal goat serum (R&D Systems) were used in permeabilized peripheral blood mononuclear cells to measure IKAROS expression.

Immunoprecipitation

HEK293T cells were co-transfected with pcDNA3-HA- IKAROS (WT or mutants) and pFlag-CMV2- IKAROS WT using Effectene transfection reagent (Qiagen). Cell lysates were prepared 20h after transfection in lysis buffer (50mM Tris [pH 7.4], 150mM NaCl, 2mM EDTA, 0.5% Triton X-100, and halt protease and phosphatase inhibitor cocktail [Thermo fisher scientific]). Total protein (500 µg) was incubated with rabbit anti-FLAG polyclonal affinity antibody (Sigma) or rabbit IgG (Cell signaling). After 2h incubation at 4 °C on a rotating wheel, 50 µl Protein A/G-agarose beads (Pierce) were added to each reaction and incubation was continued for another hour. Beads were washed three times with lysis buffer, resuspended in NuPAGE LDS sample buffer along with NuPAGE® Sample Reducing Agent (Life technology), boiled and separated on a NuPAGE® Novex® 4-12% Bis-Tris Protein Gels. Subsequent western blot analysis was performed with the mouse anti-HA monoclonal antibody (Covance), and rabbit anti-FLAG polyclonal affinity antibody.

Statistical Analysis

When indicated, data were analyzed using GraphPad Prism software (GraphPad). The differences were considered significant when $P < 0.05$.

Supplementary Figures

S1. Mutations and sequence conservation in *IKZF1*

S2. IKAROS protein structure and homology modeling

S3. *In silico* prediction of the damaging effects of the amino acid substitutions in *IKZF1*.

S4. Selective pressure on genes underlying autosomal dominant primary immunodeficiency by haploinsufficiency.

S5. IKAROS protein expression levels in T and B cells

S6. The stability of the mutant IKAROS proteins and their ability to dimerize with wild type IKAROS

S7. DNA binding and pericentromeric targeting of the mutant IKAROS proteins, EMSA and confocal microscopy

S8. Pericentromeric targeting of the mutant IKAROS proteins in HEK293T cells, confocal microscopy

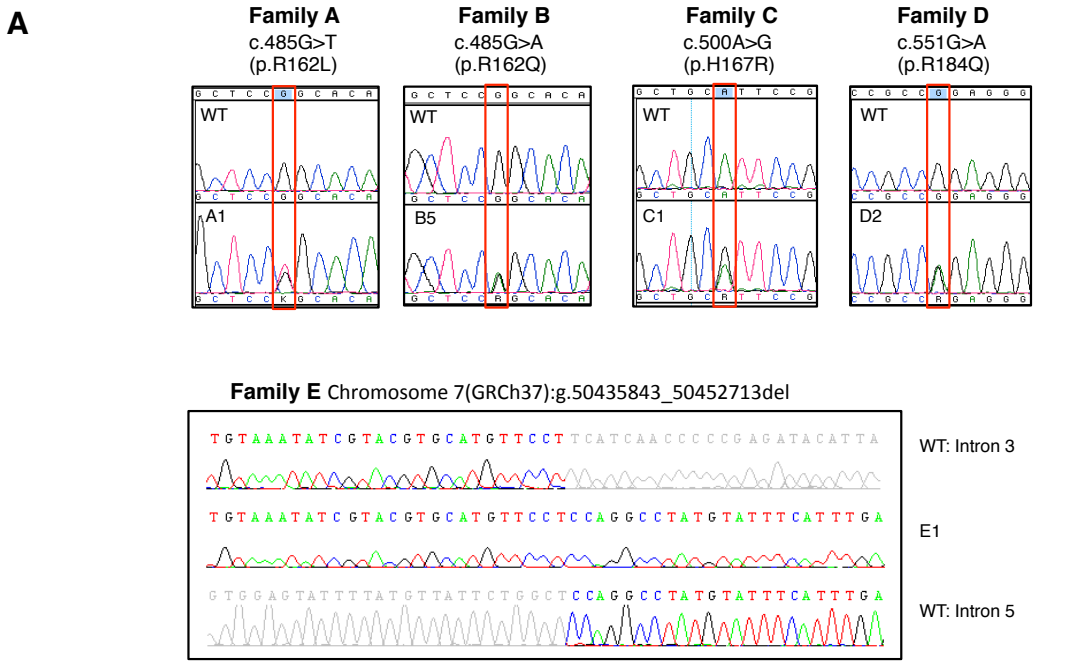
S9. Pericentromeric targeting and DNA binding of the mutant IKAROS proteins

S10. Flow cytometric TCR-Vβ repertoire analysis

S11. T cell proliferation and Fas-mediated apoptosis

S12. Penetrance of clinical and laboratory findings in patients with heterozygous mutations in *IKZF1*

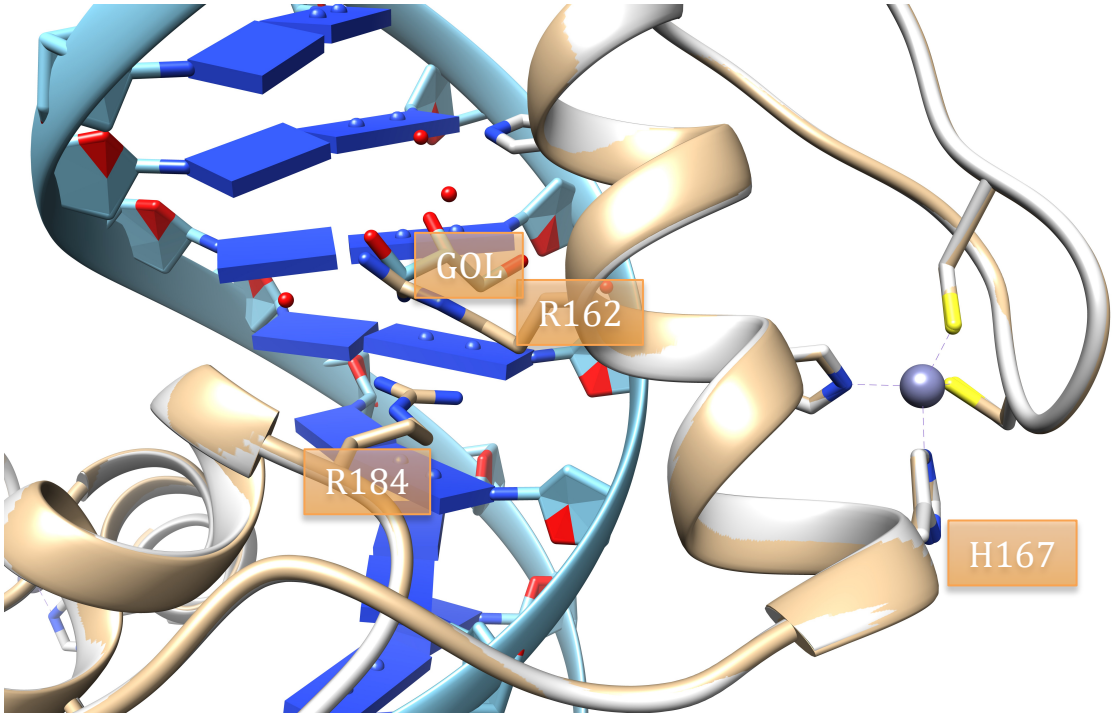
Figure S1. Mutations and sequence conservation in *IKZF1*



Panel A. Sequence chromatograms from the probands in family A-E. Chromatograms show the sequence of a healthy donor (wild type, WT) and heterozygous carriers of the mutation. Sanger sequencing over the breakpoints for the deletion in patient E1. The first part of the E1's sequence (middle sequence) matches the top sequence showing normal sequence for the region of the breakpoint in intron 3. The last part of E1's sequence matches the lower sequence showing normal sequence for the region of the breakpoint in intron 5.

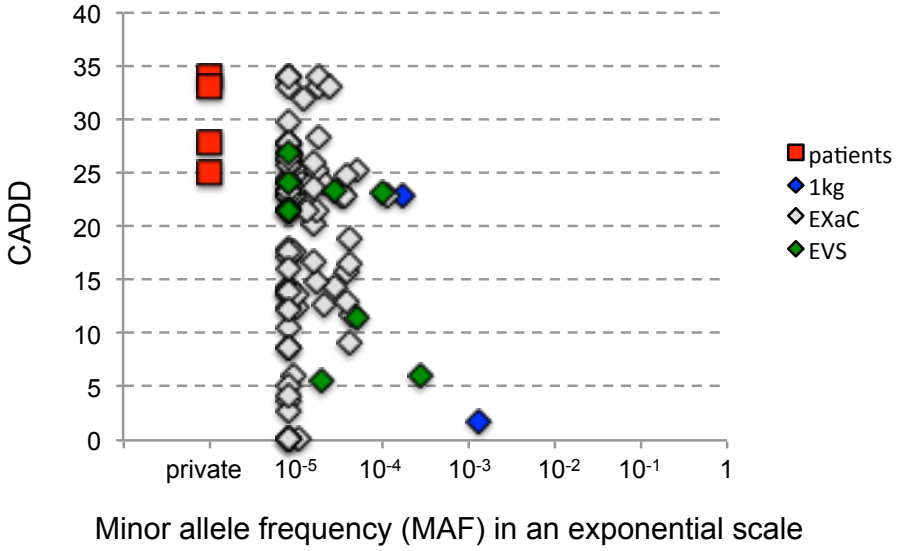
Panel B. Sequence conservation of zinc-finger domains 2 and 3 in *IKZF1*. The mutated amino acids found in Families A, B, C and D (R162L, R162Q, H167R and R184Q respectively) are indicated in red.

Figure S2. IKAROS protein structure and homology modeling



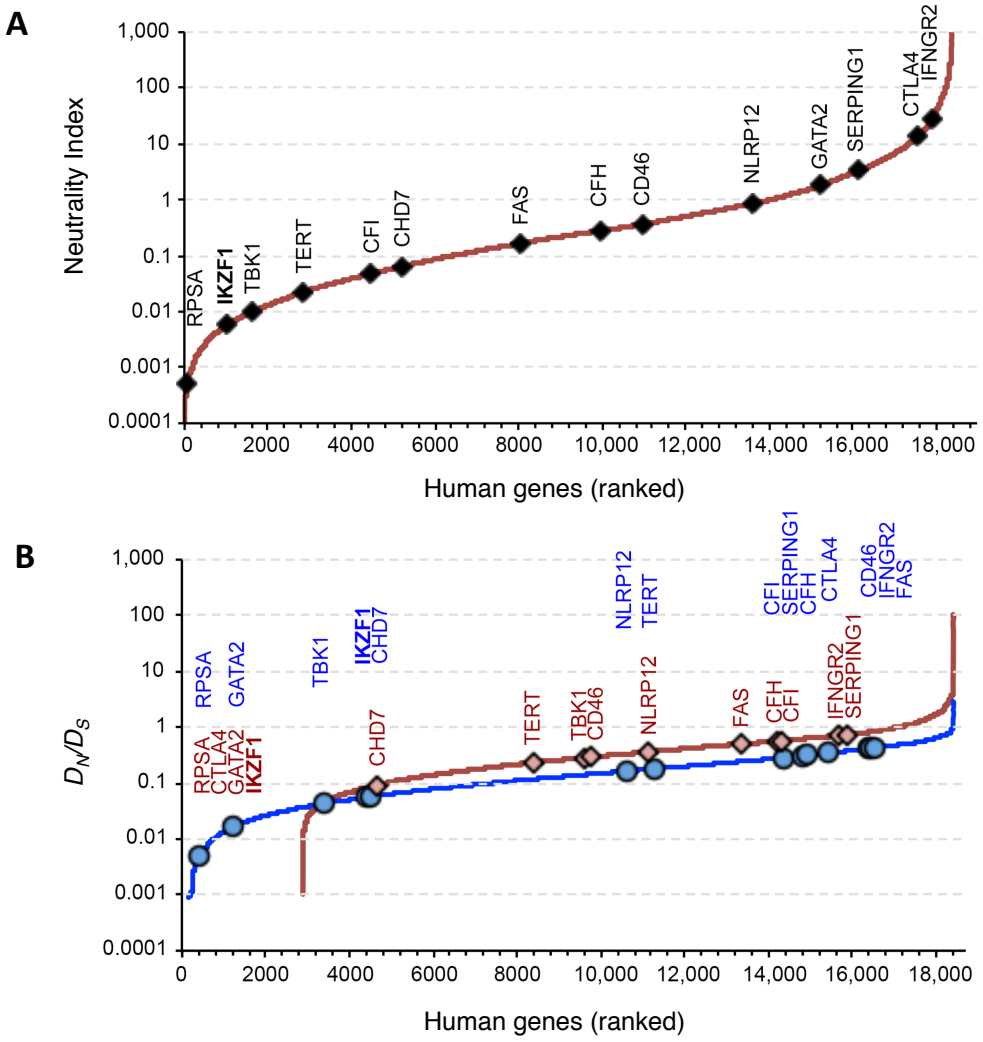
The structure of human IKAROS has not yet been determined. Therefore, to determine the approximate locations of the IKAROS variants R162L, R162Q, H167R, R184Q relative to the zinc ion coordination site and DNA binding pocket, we used UCSF Chimera[14] to superimpose a 3D protein structure homology model of wild-type human IKZF1 (UniProtKB - Q13422) from the SWISS-MODEL repository[15] onto its structural homology template, Mus musculus Aart (template)(PDB ID 2I13) bound to DNA.[16] The target-template sequence identity was 44%, indicating the potential for alignment errors in non-conserved segments of the target protein; however, the side chains of the highly conserved residues of interest do overlap with those of the corresponding template residues, suggesting accurate modeling in the regions of interest. This model suggests that the IKAROS variant H167R directly affects a zinc ion coordination site, while variants R162L, R162Q, and R184Q may affect direct or indirect (glycerol/water-mediated) binding of IKAROS to DNA due to non-conservative substitution of a positively charged arginine in the DNA binding pocket. Figure legend: Section of model of wild-type human IKAROS (tan) provided by SWISSMODEL superimposed on PDB template 2I13 (gray)[Mus musculus Aart, a six finger zinc finger designed to recognize ANN triplets] bound to DNA (cyan, blue, red). Element colors: red oxygen, blue nitrogen, yellow sulphur, purple zinc. GOL=glycerol. Water molecules are shown as red spheres.

Figure S3. *In silico* prediction of the damaging effects of the amino acid substitutions in *IKZF1*.



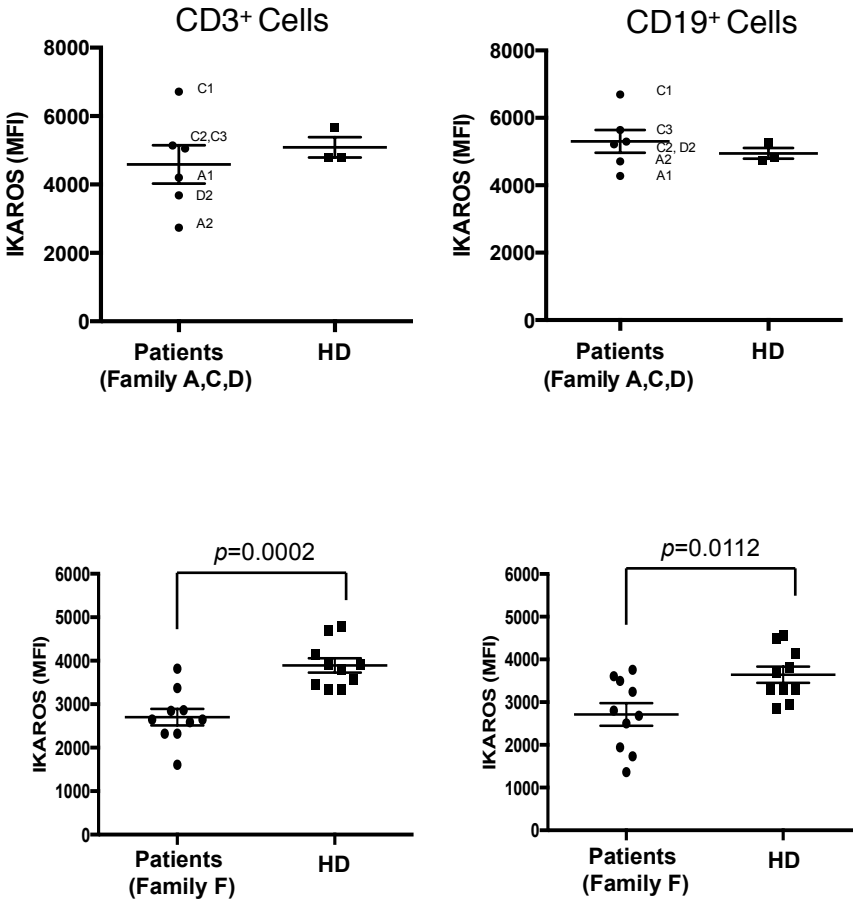
The mutations in IKAROS, R162L, R162Q, H167R and R184Q are predicted to be damaging by PolyPhen. R162L, R162L, and H167R are predicted to be deleterious and R184Q is predicted to be borderline deleterious by SIFT analysis. The Combined Annotation Dependent Depletion (CADD, v1.3) score, a method that integrates several different annotations was calculated for R162L, R162L, H167R and R184Q (red squares) and 94 variants found in the EXaC (The Exome Aggregation Consortium) database, (gray diamonds), the EVS database (green diamonds) or the 1000 genome project (blue diamonds). The ninety-four variants shown in the figure correspond to 92 missense variants, 1 in-frame deletion (minor allele frequency (MAF), 0.000008288), and 1 frameshift deletion six codons before the stop codon (Q513SfsX6; MAF, 0.000008601). The four missense mutations from unrelated patients were not found in these databases. Their MAF is therefore < 0.00001. The same applies to the deletions, not represented here.

Figure S4. Selective pressure on genes underlying autosomal dominant primary immunodeficiency by haploinsufficiency.



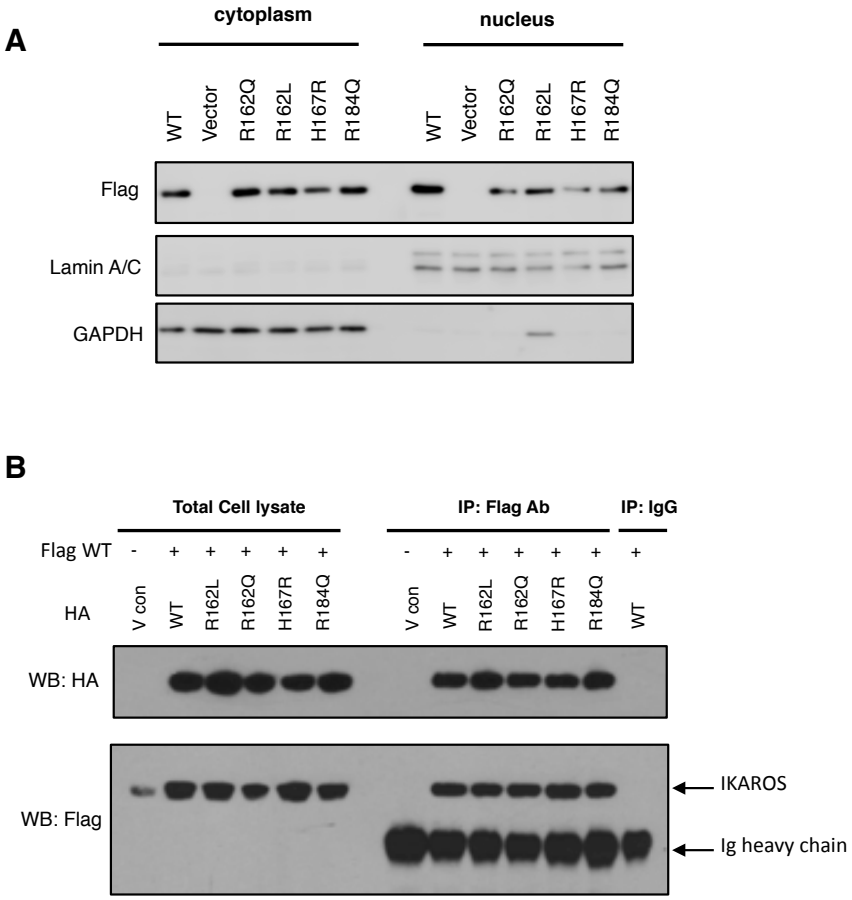
Panel A. The selective pressure acting on *IKZF1* and other genes associated with autosomal dominant immunodeficiency was determined by estimating the neutrality index (*NI*)[17,18,19] at the population level: $NI=(P_N P_S)/(D_N D_S)$, where P_N and P_S are the number of non-synonymous and synonymous alleles, respectively, at population level (1000 Genomes Project) and D_N and D_S are the number of non-synonymous and synonymous fixed sites, respectively, for the gene's coding sequence. Human genes (18,423 entries) have been ranked based on their score. Panel B. The D_N/D_S, ratio for human genes is compared with mouse (blue) or chimpanzee (red) genes respectively.[20,21]

Figure S5. IKAROS protein expression levels in T and B cells



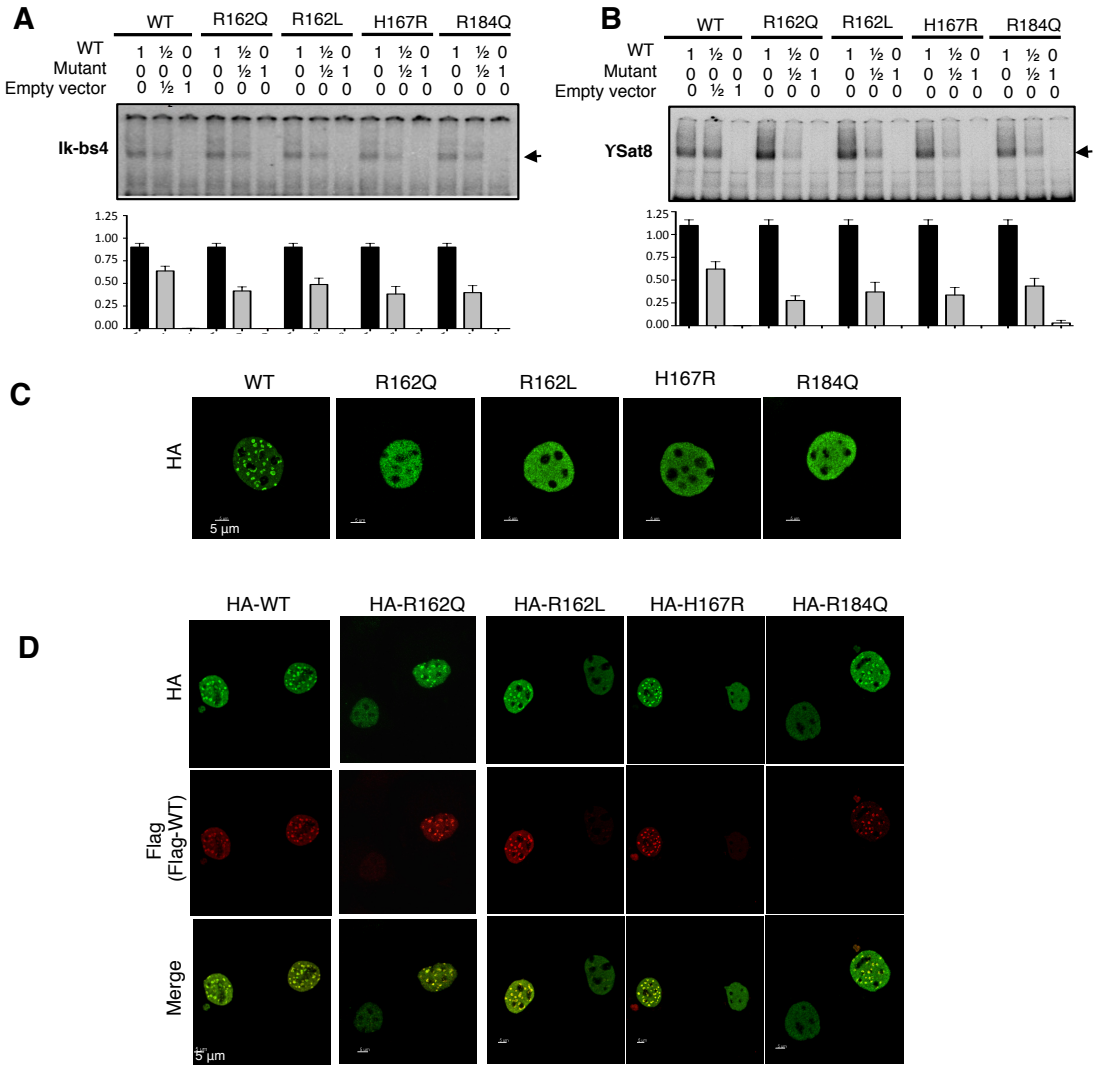
IKAROS expression levels were evaluated in affected family members and healthy donors (HD). For family D (bottom panel), ten affected members and six unaffected family members as well as four healthy donors were evaluated by flow cytometry. PE-conjugated IKAROS antibodies (for family A-C) or goat anti-IKAROS antibodies (family D) were used in permeabilized peripheral blood mononuclear cells to measure IKAROS expression (Mean fluorescence intensity [MFI]). Error bars represent the SEM; the Student's *t* test p values are shown above the bars.

Figure S6. The stability of the mutant IKAROS proteins and their ability to dimerize wild type IKAROS



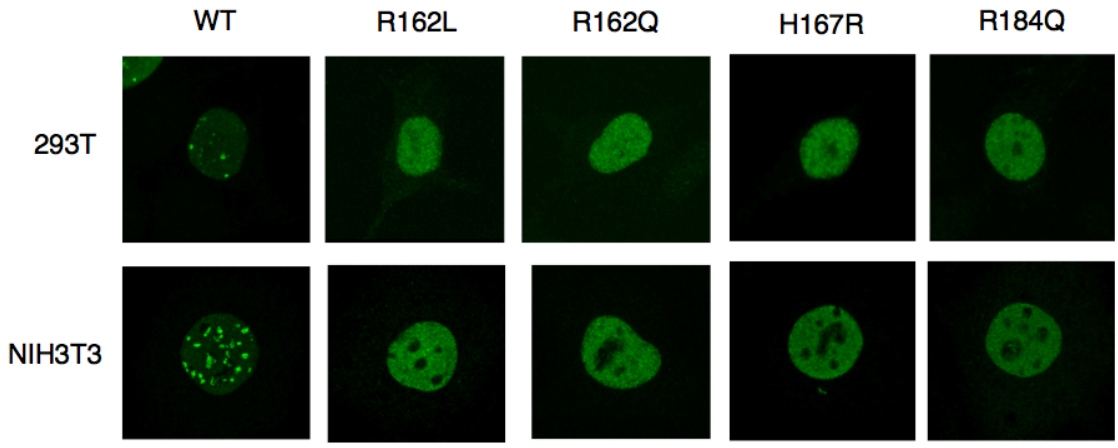
Panel A. 293T cells were transfected with Myc-tagged vectors expressing the indicated mutations and the cell lysate was divided into cytoplasmic and nuclear fractions for western blot analysis. Antibodies to Myc, Lamin A/C and GAPDH were used to document the stability of the mutant proteins, the ability of those proteins to migrate to the nucleus, and to evaluate the purity of the nuclear and cytoplasmic fractions.
Panel B. 293T cells were transfected with pFlag-CMV2-IKAROS wild type (WT) and pcDNA3-HA-IKAROS WT or mutants (R162L, R162Q, H167R, and R184Q). V indicates empty vector control. Immunoprecipitations were performed using anti-Flag antibodies or rabbit IgG. Western blot analysis of the IP samples with anti-HA (upper panels) and anti-Flag antibodies (lower panels) is shown. Total cell lysates was loaded with 5% of the protein extract used for each IP reaction. Data shown are representative of 3 experiments.

Figure S7. DNA binding and pericentromeric targeting of the mutant IKAROS proteins, EMSA and confocal microscopy



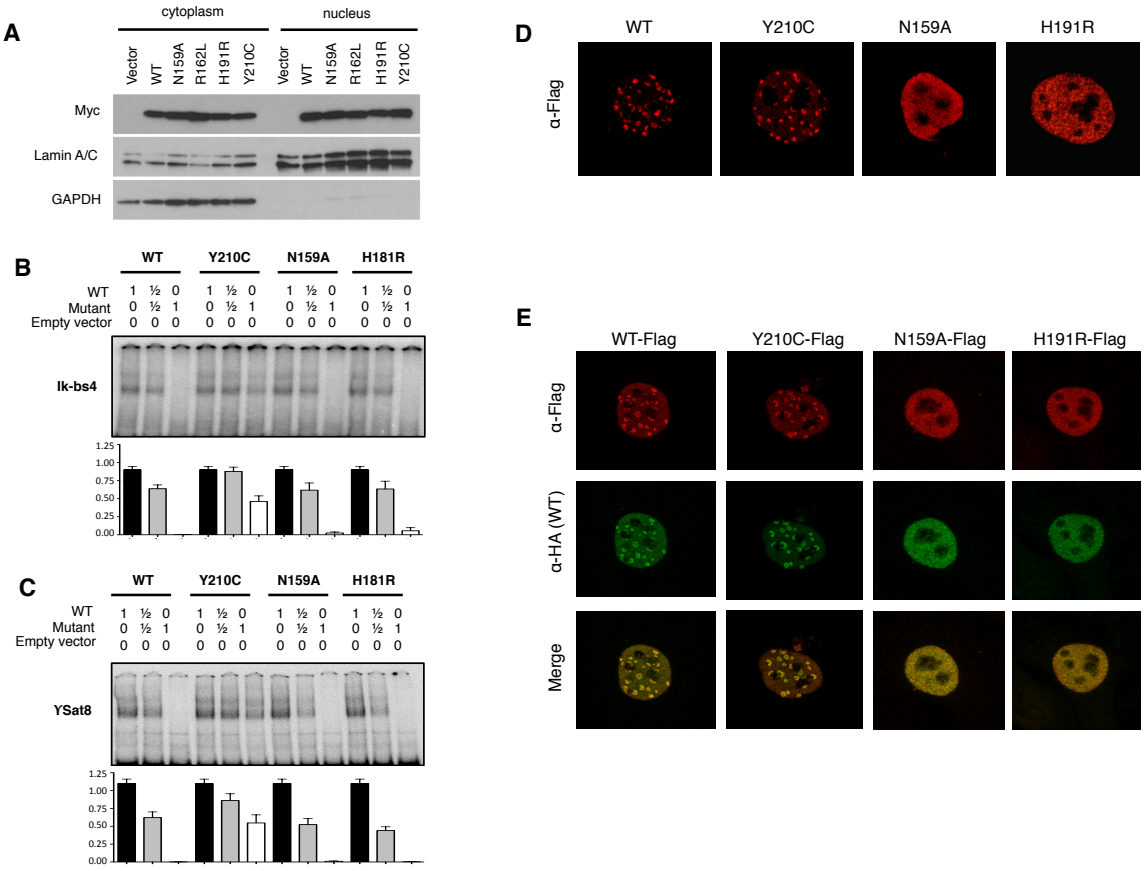
A, B. EMSAs were performed using nuclear extracts from HEK293T cells transfected with 500 ng of vector expressing WT IKAROS, 250 ng of WT plus 250 ng of mutant IKAROS expression vectors, or 500 ng of mutant IKAROS expression vector. The extracts were allowed to bind to 2 different IKAROS probes: IK-bs4, an IKAROS consensus binding sequence and γ Sat 8, a sequence from the pericentromeric region of human Chromosome 8. IKAROS containing complexes are indicated with an arrow. Data are normalized using the binding of the extracts from the WT transfected cells as 100%. Errors bars represent the SEM (2 or 3 independent experiments). C. NIH3T3 cells were transfected with HA-tagged WT or mutant expression vectors and were labeled with anti-HA antibody and an Alexa 488-conjugated (green) secondary antibody. Cells were visualized by using a confocal microscope. D. NIH3T3 cells were co-transfected with HA-tagged WT or mutant vectors (green), plus FLAG-tagged vectors expressing WT IKAROS and developed using an anti-FLAG antibody followed by Alexa 568-conjugated (red) antibody. Cells that received both the mutant and WT vectors demonstrated the typical punctate staining. Data shown are representative of 3 confocal experiments.

Figure S8. Pericentromeric targeting of the mutant IKAROS proteins in HEK293T cells



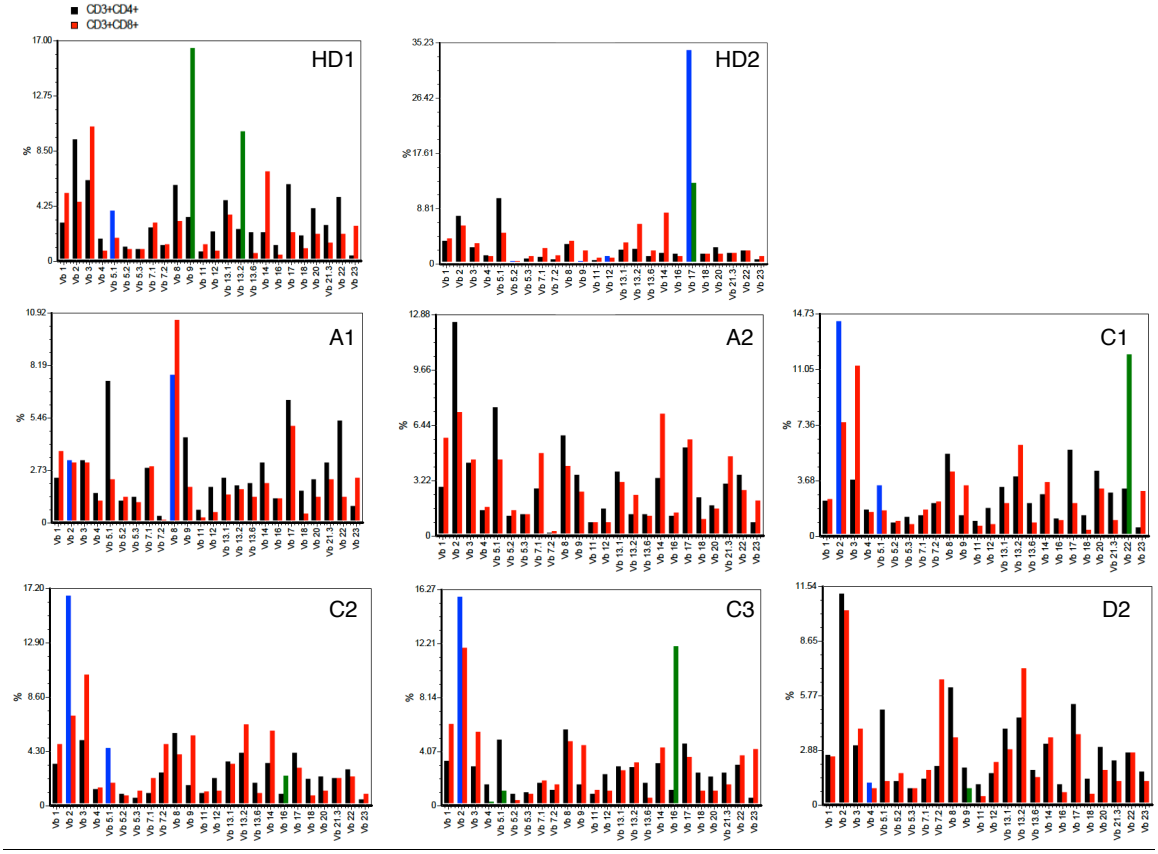
Human embryo kidney (HEK) 293T cells were transfected with pCMV6-AC-Myc-DDK-IKAROS WT or indicated mutant. After 20 hours, cells were stained with a Flag antibody, followed by Alexa Fluor 488 antibody. Pictures were acquired using identical confocal settings (results on similar experiments done on mouse NIH3T3 cells are included for comparison).

Figure S9. Pericentromeric targeting and DNA binding of the mutant IKAROS proteins



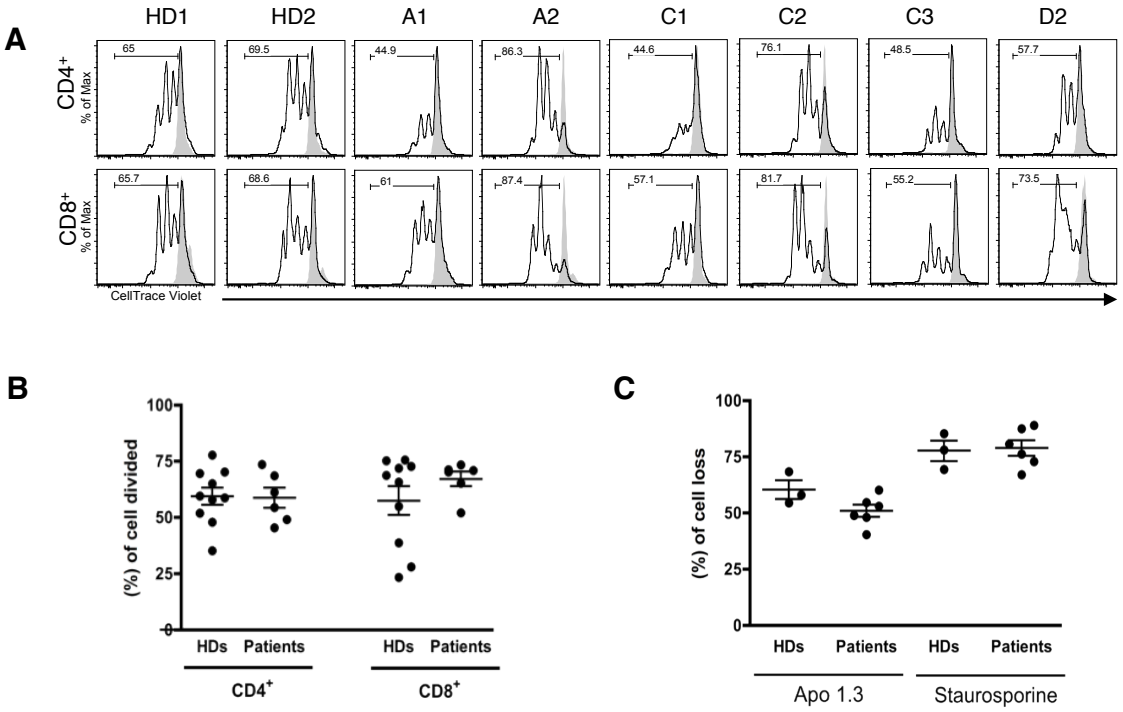
Panel A. 293T cells were transfected with pCMV6-AC-Myc-DDK-wild type or mutants. Western blot data shows the relative amount of each protein. Panels B and C. EMSAs were performed with nuclear extracts using 2 different probe for IKAROS binding sites. Data are normalized using the wild type transfected cells as 1. Errors bars represent the SEM (2 or 3 independent experiments). Panel D, The confocal images of NIH3T3 cells transfected with indicated mutant vector. In all images, cells were stained with a Flag antibody, followed by Alexa Fluor 568 antibody. E. NIH3T3 cells were co-transfected with pCMV6-AC-Myc-DDK-wild type or mutants, plus HA-tagged vectors expressing wild type IKAROS. In all images, cells were stained with an anti-HA antibody and an anti-Flag antibody, followed by fluorochrome-conjugated secondary antibodies. Data shown are representative of 3 confocal experiments.

Figure S10. Flow cytometric TCR-Vβ repertoire analysis



Whole blood samples from two healthy donors (HD) or 6 patients were stained with fluorochrome-conjugated CD3, CD4, CD8, and TCR-VB families antibodies according to manufacturer's instruction (Beckman Coulter; IOTest® Beta Mark TCR V beta Repertoire Kit). Cells were analyzed by flow cytometry (Becton Dickinson FACSCanto II). Blue indicates out of range for CD4 T cells, green is out of range for CD8 T cells

Figure S11. T cell proliferation and Fas-mediated apoptosis



Panel A. Peripheral blood mononuclear cells were stained with CellTrace Violet and stimulated with (black line) or without (solid gray) anti-CD3 and anti-CD28 antibodies (1µg/mL) for 3 days. Numbers indicate percentage of cells having undergone at least one cellular division, assessed by dye dilution. Proliferation data shown are representative of 3 experiments. Panel B. Graph indicates percentages of cells divided from healthy donors and patients. Data are means ± SEM of replicates of 6 patients and 10 healthy donors. Panel C. Total PBMCs were stimulated with anti-CD3 and CD28 for 3 days, and then cultured cells with IL-2. At day 11, cells were stimulated with Apo1.3 in the presence of protein A, or staurosporine for 20 hours. Cells were stained with PI and (%) of cell loss are calculated as described in the method. Data are means ± SEM from 3 healthy donors and 6 patients.

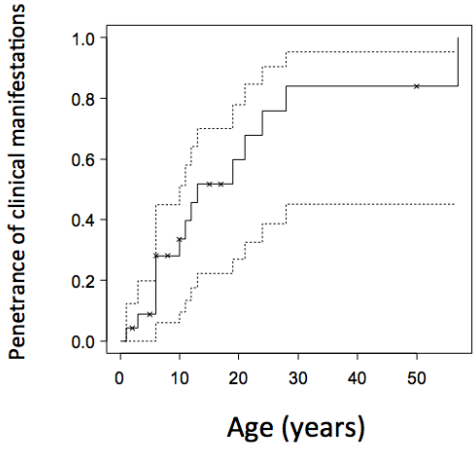
Figure S12. Penetrance of clinical and laboratory findings in patients with heterozygous mutations in *IKZF1*

A

Patient	Clinical		Immunological	
	Manisfestations	Age at onset (years)	Serum IgG, IgA, and/or IgM	B cells
A1*	INFECTION	9	Abnormal	LOW
A2	ITP INFECTION	3 6	Abnormal	LOW
B1†	INFECTION	12	Abnormal	not available
B2	INFECTION	6	Abnormal	LOW
B3	INFECTION	10	Abnormal	Normal
B4	INFECTION	6	Abnormal	LOW
B5*	INFECTION	3	Abnormal	LOW
B6	INFECTION	1	Abnormal	LOW
B7†	B-ALL	3	not available	not available
C1*	INFECTION	32	Abnormal	LOW
C2	INFECTION	13	Abnormal	LOW
C3	INFECTION	11	Abnormal	LOW
D1	ASYMPTOMATIC		Abnormal	Normal
D2*	INFECTION	9	Abnormal	LOW
E1*	INFECTION	3	Abnormal	LOW
E2	ASYMPTOMATIC		not available	not available
F1	INFECTION	57	Abnormal	LOW
F2*	INFECTION	29	Abnormal	LOW
F3	INFECTION	21	Abnormal	LOW
F4	INFECTION	24	Abnormal	LOW
F5	INFECTION	28	Abnormal	LOW
F6	INFECTION	19	Abnormal	LOW
F7	ASYMPTOMATIC		Abnormal	LOW
F8	ASYMPTOMATIC		Normal	LOW
F9	ASYMPTOMATIC		Abnormal	LOW
F10	ASYMPTOMATIC		Abnormal	Normal
F11	ASYMPTOMATIC		Abnormal	Normal
F12	B-ALL INFECTION	5 6	Abnormal	Normal
F13	ASYMPTOMATIC		Abnormal	Normal
Penetrance : Abnormal/total % penetrance	**15/23 65%**		**20/21 95%**	**14/20 70%**

* : index cases, †: dead form infection or leukemia

B



We estimated clinical penetrance by excluding index cases, focusing on the 15 symptomatic and 8 asymptomatic genetically affected family members. The overall penetrance of clinical manifestations was found to increase progressively over time. x indicates the age of asymptomatic family members at the time of the most recent evaluation. Penetrance curves as a function of age were estimated by the Kaplan-Meier method, using the survival package of the R software v3.0.2 (http://cran.r-project.org/). The small number of patients accounts for the large confidence intervals.

Supplementary tables

S1. Clinical and laboratory features in patients with heterozygous IKAROS mutations

S2. List of variants identified among the list of 269 genes associated with PID in the patients evaluated by WES.

S3. TdT mediated N nucleotide additions within VH3 heavy chain CDR3 sequences

Table S1. Clinical and laboratory features in patients with heterozygous IKAROS mutations

Pt	Clinical manifestations (age at onset in years)	Age at study in years	IgG	IgA	IgM	CD19 B cell %(ALC)	CD27+ among CD19+ %(ALC)[&]	CD3 T cell %(ALC)	CD3+CD4+ %(ALC)	CD3+CD8+ %(ALC)	CD4/CD8 Ratio	CD3- CD16+/56+ NK cell %(ALC)	Vaccine titers (age at study in years)
A1	Infections (9)	10	317	<10	<10	0.4(15)	-	94.5(3536)	30.2 (1167)	55.2(2133)	0.55	-	PPV[-] [0/4],
		20	<33	<7	21	0.5(17)	-	95.3(3180)	33.8(1128)	59.4(1982)	0.57	3.4(113)	TT[-], Hib[-] (10)
		29	-	-	-	0.1(4)	-	97.5(3851)	39.1(1544)	54(2133)	0.72	2.3(91)	
A2	ITP (3)	5	434	-	-	-	-	-	-	-	-	-	
	Infections (6)	6	727*	7	5	0.3(12)	-	85.5(3129)	47.2(1831)	34(1319)	1.39	12.3(477)	
		11	-	-	-	0.2(8)	-	86.1(3539)	42(1726)	38(1562)	1.11	13.4(551)	
B1	Infections (12)#	71	386	30	60	-	-	-	-	-	-	-	
B2	Infections (6)	47	461	36	45	3(80)	15(12)	85(2350)	34.5(950)	51(1410)	0.67	11(300)	TT[-] (47)
B3	Infections (10)	46	283	15	71	7(260)	72(187)	91(3330)	31.2(1140)	61(2230)	0.51	3(110)	TT[-], DT[-], Hib[+] (46)
B4	Infections (43)	43	433	19	56	2(50)	45(22)	86(2030)	22.9(540)	65(1530)	0.35	11(260)	TT[-], DT[-], Hib[+] (43)
B5	Infections (3)	23	1430*	21	22	2(50)	3(1)	76(1900)	36(900)	38(950)	0.95	22(550)	
B6	Infections (1)	17	1260*	12	21	1(20)	17(3.4)	79(1560)	38(750)	36(710)	1.05	21(420)	
B7	B-ALL (3)#	3	-	-	-	-	-	-	-	-	-	-	
C1	Infections (30)	32	315	12	30	-	-	-	-	-	-	-	
		34	1280*	35	21	0.8(33)	-	88(3656)	27(1122)	53.8(2235)	0.5	11.5(478)	
		43	821*	97	<21	1(40)	-	84.1(3451)	29.4(1203)	49.8(2039)	0.59	14.2(583)	
		46	887*	73	11	1(26)	-	83.4(2177)	30.9(806)	49.8(1300)	0.62	15.6(407)	
C2	Infections (13)	0.8	160	<7	13	3.2(287)	-	87.8(7918)	50.4(4543)	37.7(3403)	1.34	9(809)	TT[-], DT[-],
		3	1000*	73	17	1.5(49)	-	90(2916)	50.7(1643)	33.9(1098)	1.36	8.3(269)	VZV[-], Rub[-] (13)
		13	761	136	<21	0.7(18)	-	85.6(2343)	33.4(928)	44.9(1247)	0.73	12.5(336)	
		16	1052*	157	12	0.2(5)	-	83.3(2049)	25.5(627)	49.8(1225)	0.51	16.2(399)	
C3	Infections (4)	1.5	328	15	6	7.8(432)	-	84.8(4877)	50.2(2968)	31.9(2052)	1.57	3.9(196)	TT[+], DT[+] (1.5)
		4	308	<7	5	4.9(307)	-	86(5389)	48(3052)	39(2450)	1.2	9(589)	TT[-], DT[-] (4)
		11	760*	22	<21	1.3(53)	15(8)	82.2(3273)	36.7(1460)	40.2(1600)	0.91	16.5(659)	
		14	1030*	12	<6	0.4(18)	-	82.3(3712)	25.9(1168)	47.6(2147)	0.54	17.7(798)	
D1	Asymptomatic	50	637	5	45	7(160)	-	89(2022)	29(632)	53(1397)	0.45	3(74)	TT[+], DT[+] (50)
D2	Infections (9)	9	89	5	7	-	-	-	-	-	-	-	
		23	42	<1	7	0.3(9)	-	96(3388)	40(1416)	54(1888)	0.75	4(128)	
E1	Infections (3)	4	460	<7	33	2(84)	-	67(2930)	30 (1312)	30(1312)	1	-	
		11	820*	<6	11	0.8(29)	-	83.3(2274)	40(1092)	21(573)	1.9	10.3(281)	
		29	860*	<1	10	0.2(5)	-	93(2817)	25(787)	67(1966)	0.4	4(123)	
E2	Asymptomatic	6	-	-	-	-	-	-	-	-	-	-	
F1	Infections (57)	57	211	81	19	0.8(7)	-	77.6(698)	41.1(287)	35.8(250)	1.15	20.1(181)	
		62	113	<6	13.9	0.8(11)	22.5(2)	66(866)	26 (343)	38(498)	0.7	32(423)	
F2	Infections (29)	29	<33	<7	<5	-	-	-	-	-	-	-	PPV[-] [0/14] (57)
		57	<7	<4	<2	1.0(18)	-	94(1709)	59(1068)	35(640)	1.69	-	
		66	-	-	-	0.3(7)	-	84.2(1768)	-	-	-	16.8(353)	
F3	Infections (21)	21	79	<7	14	1.0(15)	-	86(1268)	44(649)	42(619)	1.05	13(188)	
		34	-	-	-	0.8(9)	-	79.2(950)	60.1(571)	39.9(379)	1.51	20.2(242)	
		37	1010*	<6	4.6	0.6(8)	23.7(2)	79(1020)	40.4(522)	37.6(486)	1	19.7(254)	
F4	Infections (22)	24	121	<4	<2	-	-	-	-	-	-	-	
		25	120	<7	<10	1.0(14)	-	82(1229)	45(660)	37(537)	1.22	15(233)	
		31	-	-	-	0.2(1)	-	93.5(1683)	36.7(618)	59(993)	0.62	6.6(119)	
		36	1010*	<6	6.6	0.1(2)	57.9(1)	95.3(2489)	31.1(787)	64.6(1686)	0.4	4.2(111)	
F5	Infections (31)	28	391	33	44	0.3(6)	-	92.6(1945)	51(1005)	50(990)	1.02	7.1(149)	
		31	382	29	21	0.5(12)	19.3(2)	87.3(2078)	34.7(822)	52.2(1238)	0.66	11(260)	
F6	Infections (19)	42	395	16	15	-	-	-	-	-	-	-	
		47	148	32	39	5.5(138)	3.9(5)	82.2(2115)	47.8(1231)	34.4(886)	1.39	11.7(302)	
F7	Asymptomatic	8	659	1	11	0.3(6)	-	90.4(1627)	54.2(882)	30.7(500)	1.77	9.2(166)	
		12	528	<6	18	0.5(9)	17.8(2)	84.7(1464)	57.6(996)	27.1(469)	2.12	13.5(233)	
F8	Asymptomatic	5	789	64	46	2.7(89)	-	86.4(2851)	53.5(1525)	31.8(907)	1.68	7.4(244)	
F9	Asymptomatic	2	793	31	50	7.4(221)	-	77.7(2331)	29.4(685)	47.4(1105)	0.62	12.8(384)	
		6	844	28	36	15(354)	6.3(22)	77.1(1940)	34.4(866)	39.6(955)	0.87	7.1(179)	
F10	Asymptomatic	17	514	71	40	-	-	-	-	-	-	-	
		21	606	118	32	7.5(187)	22.4(42)	76.2(1864)	43.7(1068)	31(758)	1.41	15.4(377)	
F11	Asymptomatic	15	490	70	32	-	-	-	-	-	-	-	
		20	541	120	53	17.1(544)	14.6(79)	68(2026)	43.1(1286)	22(656)	1.96	13.5(403)	
F12	B-ALL (5)	13	451	32	38	-	-	-	-	-	-	-	
	Infections (6)	17	590	36	12.6	13.7(452)	9.2(42)	75.1(2292)	33.8(1032)	41.6(1269)	0.81	9.8(299)	
F13	Asymptomatic	10	505	19	24	-	-	-	-	-	-	-	
		14	629	28	40	15.4(291)	11.8(34)	77.7(1356)	54.6(953)	21.7(378)	2.52	6.3(110)	

Numbers in Red/Blue indicate values above/below the normal range (respectively) compared to age-matched controls in the laboratory in which the study was performed: NIH Clinical Center (Family A, C), University Children's Hospital Zurich (Family B), Quest Diagnostics (Family D), Oslo University Hospital and University of Oslo, Norway (Family E), and ARUP Laboratories (Family F). Serum IgG, IgA and IgM are expressed in mg/dL; *Patient was receiving Immunoglobulin (IgG) replacement therapy; ALC, absolute lymphocyte count; [&]Values for $CD27^+$ B cells are included only if at least 50 $CD19^+$ or $CD20^+$ B cells were analyzed. Vaccine titers: [+] Protective, [-] Non-protective, for pneumococcal vaccine, [serotype-specific protective titers/serotypes tested are depicted in parenthesis]; PPV, pneumococcal polysaccharides vaccine; TT, Tetanus toxoid; DT, Diphtheria Toxoid; Hib, *Haemophilus influenzae* type b vaccine; VZV, Varicella-zoster vaccine; Rub, rubella vaccine; Infect, Infections; ITP, Idiopathic thrombocytopenic purpura; B-ALL, B-cell Acute Lymphoblastic Leukemia; # Deceased patient

Table S2. List of 269 genes associated with PID that were screened in the patients analyzed by WES.

ACP5, ADA, ADAR, AICDA, AIRE, AK2, AP3B1, AP3D1, APOL1, ATM, B2M, BCL10, BLM, BLNK, BLOC1S6, BTK, C1QA, C1QB, C1QC, C1R, C1S, C2, C3, C4A, C4B, C5, C6, C7, C8A, C8B, C8G, C9, CARD11, CARD14, CARD9, CASP10, CASP8, CCBE1, CCBEB1, CD19, CD247, CD27, CD3D, CD3E, CD3G, CD40, CD40LG, CD46, CD59, CD70, CD79A, CD79B, CD81, CD8A, CEBPE, CFB, CFD, CFH, CFHR1, CFI, CFP, CIB1, CIITA, CLPB, COLEC11, COPA, CORO1A, CR2, CSF2RA, CSF2RB, CTLA4, CTPS1, CXCR4, CYBA, CYBB, DCLRE1B, DCLRE1C, DKC1, DNA ligase 1, DNMT3B, DOCK2, DOCK8, ELANE, EP45, FADD, FAS, FASLG, FCN3, FERMT3, FOXN1, FOXP3, G6PC3, G6PD, GATA2, GFI1, GINS1, HAX1, IFIH1, IFNGR1, IFNGR2, IGHM, IGKC, IGLL1, IKBKB, IKBKG, IKZF1, IL10, IL10RA, IL10RB, IL12B, IL12RB1, IL17F, IL17RA, IL17RC, IL18, IL1RN, IL21, IL21R, IL2RA, IL2RG, IL36RN, IL7R, INO80, IRAK4, IRF3, IRF4, IRF7, IRF8, ISG15, ITCH, ITGAX, ITGB2, ITK, JAGN1, JAK3, KRAS, LAMTOR2, LCK, LIG4, LIPA, LPIN2, LRBA, LYST, MAGT1, MALT1, MAP3K14, MASP1, MASP2, MBL2, MCM4, MEFV, MKL1, MOGS, MPO, MRE11A, MSN, MVK, MYD88, NBN, NCF1, NCF2, NCF4, NFAT5, NFKB1, NFKB2, NHEJ1, NHP2, NLRC4, NLRP12, NLRP3, NOD2, NOP10, NRAS, ORAI1, PCNA, PGM3, PIK3CD, PIK3R1, PLCG2, PMS2, PNP, POLE, PRF1, PRKCD, PRKDC, PSMB8, PSTPIP1, PTPN6, PTPRC, RAB27A, RAC1, RAC2, RAG1, RAG2, RBCK1, RFX5, RFXANK, RFXAP, RHOH, RLTPR, RMRP, RNASEH2A, RNASEH2C, RNASEL, RNF168, RNF31, RORC, RPSA, RTEL1, SAMHD1, SBDS, SERPING1, SH2D1A, SLC35C1, SMARCAL1, SP110, SPINK5, SPPL2A, STAT1, STAT2, STAT3, STAT5B, STIM1, STK4, STX11, STXBP2, TADA2A, TAP1, TAP2, TAPBP, TBK1, TBX1, TCC37, TCF3, TECR, TERC, TERT, TGFBR1, TGFBR2, TICAM1, TINF2, TLR3, TMC6, TMC8, TMEM173, TNFRSF13B, TNFRSF13C, TNFRSF1A, TNFRSF4, TNFSF12, TPP2 , TRAC, TRAF3, TRAF3IP2, TREX1, TTC37, TTC7A, TYK2, UNC119, UNC13D, UNC93B1, UNG, USB1, VPS45, WAS, WIPF1, XIAP, ZAP70, ZBTB24).

Table S3. TdT mediated N nucleotide additions within VH3 heavy chain CDR3 sequences

TdT mediated N nucleotide additions within VH3 heavy chain CDR3 sequences

Subject	Unique Sequences	N1	N2	Total N	CDR3 Length
Control 1	20	5.7	4.8	10.6	15.6
Control 2	12	5.8	7.3	13.0	14.3
Control 3	16	7.5	6.5	14.0	15.2
Control 4	12	5.5	7.1	12.6	16.1
BTK R255X	8	6.3	3.7	10.0	13.8
BTK L452P	10	4.4	4.8	9.3	14.4
IKZF1 H167R	32	5.9	8.5	14.4	16.5
IKZF1 R162Q	21	5.3	4.1	9.4	15.9

The IMGT/V-QUEST website tool was used for sequence analysis. N1 and N2 indicate the average number of nucleotide additions between V and D regions and between the D and J regions respectively. The length of the CDR3 region in amino acids is shown.

References

1. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
2. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.
3. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.
4. Bayram Y, Aydin H, Gambin T, et al. Exome sequencing identifies a homozygous C5orf42 variant in a Turkish kindred with oral-facial-digital syndrome type VI. Am J Med Genet A 2015;167:2132-7.
5. Reid JG, Carroll A, Veeraraghavan N, et al. Launching genomics into the cloud: deployment of Mercury, a next generation sequence analysis pipeline. BMC bioinformatics 2014;15:30.
6. Challis D, Yu J, Evani US, et al. An integrative variant analysis suite for whole exome next-generation sequencing data. BMC bioinformatics 2012;13:8.
7. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
8. Samarakoon PS, Sorte HS, Kristiansen BE, et al. Identification of copy number variants from exome sequence data. BMC genomics 2014;15:661.
9. Kastner P, Dupuis A, Gaub MP, Herbrecht R, Lutz P, Chan S. Function of Ikaros as a tumor suppressor in B cell acute lymphoblastic leukemia. Am J Blood Res 2013;3:1-13.
10. Sellars M, Kastner P, Chan S. Ikaros in B cell development and function. World J Biol Chem 2011;2:132-9.
11. Monroe JG, Rothenberg EV. Molecular biology of B-cell and T-cell development. Totowa, N.J.: Humana Press; 1998.
12. Cobb BS, Morales-Alcelay S, Kleiger G, Brown KE, Fisher AG, Smale ST. Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding. Genes Dev 2000;14:2146-60.
13. Ganapathi KA, Townsley DM, Hsu AP, et al. GATA2 deficiency-associated bone marrow disorder differs from idiopathic aplastic anemia. Blood 2015;125:56-70.
14. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605-12.
15. Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T. The SWISS-MODEL Repository and associated resources. Nucleic Acids Res 2009;37:D387-92.
16. Segal DJ, Crotty JW, Bhakta MS, Barbas CF, 3rd, Horton NC. Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA. J Mol Biol 2006;363:405-21.
17. Rand DM, Kann LM. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol Biol Evol 1996;13:735-48.
18. Cock PJ, Antao T, Chang JT, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 2009;25:1422-3.

19. Stoletzki N, Eyre-Walker A. Estimation of the neutrality index. Mol Biol Evol 2011;28:63-70.
20. Kimura M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 1977;267:275-6.
21. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. Trends Ecol Evol 2000;15:496-503.